Opening up the archives of the British Antarctic Survey

BEVERLEY AGER
British Antarctic Survey,
Cambridge, United Kingdom

Abstract

The Information Services group at the British Antarctic Survey (BAS) – which includes the Polar Data Centre and Archives Service – aims to promote the visibility and availability of its collections. This includes making improvements to metadata catalogue descriptions, and increasing access to polar data and information. This presentation will provide an overview of a recent project undertaken by the Archives Service, aiming to increase access to its holdings, improve data exchange with other systems and provide open linked data.

The Archives collection reflects over 70 years of interdisciplinary science in the polar regions with roughly 1.2 linear km physical records and 9Tb digital data. The initial part of the project involved migration of 40,000 catalogue records to xml-based Modes Complete software. This software is often used for museum collections and combines high-functionality with cost-effectiveness. It provides tools for managing mixed collections in a single cataloguing system that complies with a variety of archival descriptive standards.

Currently the focus of the project is on working to make the catalogue available online using CKAN open source software. This will provide discovery-level catalogue descriptions and access to images and digital versions of documents and films. This portal will also be used cross-BAS to provide access to polar collections, including the datasets held by the Polar Data Centre, the UK's national repository for polar data.

Introduction

The Information Services group at the British Antarctic Survey (BAS) consists of the Library, Polar Data Centre, Archives Service and Web & Apps teams. Jointly, these teams are responsible for the management of seven decades' worth of polar data and information. The group's core remit is to promote the visibility and availability of its collections through improvements to metadata catalogue descriptions and data access systems. This paper will focus on a current project to develop a public web interface for the archives database, covering the background to the project and future developments, including interaction with other systems to create linked networks of polar data and information.

Background

Collections overview: a mixed archive

The scope and extent of the BAS Archives reflect over seventy years of interdisciplinary science in the Polar Regions, making it a unique resource of scientific activities and administrative and logistical operations in the Arctic and Antarctic. There is a roughly 50:50 ratio of administrative records to science data, and about 10% of the collection is material donated by past members of staff. In total, this amounts to some 1.2 linear km of physical records and 9Tb of digital data. The collection includes semi-current business records, artworks, artefacts, and oral history recordings, and covers a wide variety of formats – written, photographic, printed maps and books, film, sound recordings and digital data. Importantly, the BAS Archives is also an approved Place of Deposit, as assigned by The National Archives, and so preserves and provides access to public records. BAS is also subject to The Freedom of Information Act, under which the public can request information relating to organisational records held, and so there is a legal obligation to be open and accessible.

Demonstrating the mixed/multi-faceted nature of the Archives, the above image shows surveyor, Robin Sherman, plane-tabling on Signy Island, South Orkney Islands, Jan 1958. The collection also includes items of clothing from the period, surveying equipment, panarounds, and survey reports. The final map produced from this surveying trip is also held in the Archives. BAS Archives ref: G21/1/A431/7 © Sherman, Robin Lewis

Archives software requirements and functionality

The imminent completion of the web interface will mark the culmination of a long-running project to expose the Archive collections online. This began with the need to update the old archives database, based on Modes for Windows software. This software had been superseded by a new version so there were no further upgrades available, plus it lacked compatibility with newer computer hardware and platforms. A key driver was also the need to upgrade to a product that would provide increased functionality with the potential to develop online access and improved data-sharing. The migration phase of the project, led by former BAS archivist, Joanna Rae, involved an extensive exercise to map fields across to the new software, as well as the comprehensive cleaning of thousands of records. The catalogue itself was over 30 years' old and had accumulated cataloguing, or rather, cataloguer inconsistencies, such as idiosyncratic field usage or non-standard text formatting. The migration was completed in 2016 with nearly 60,000 archive catalogue entries in total migrated to a new version of the software, Modes Complete.

The use of Modes software is more established in museum rather than archive environments but was selected for its high functionality – delivered at a fraction of the cost of traditional archives packages. It is also used by the Scott Polar Research Institute, and the archivists at BAS and SPRI were able to work closely with Modes developers to expand the software for archival specific use. This has allowed archivists to use it more straightforwardly to meet the needs of their own descriptive standards. The resulting product is compliant with a variety of standards and schema (ISAD(G), EAD, ISO19115, Spectrum).

The BAS Archives' iteration of Modes uses the standard software framework to better integrate future upgrades and allow for easier onward migrations. However, it also has the flexibility to accommodate in-house customisation and internal cataloguing conventions. This means that concepts not specifically catered for in the data structure can be modified for BAS-specific use by adding 'attributes' to elements or employing 'type' elements to define usage. It also allows access to external and in-house term lists and validation rules to standardise data entry. Another factor is its ability to cope with very long records as the Archives have been catalogued to a high-level of granularity. This level of detail has benefits for the repurposing of content, for example, as contextual metadata for other data portals (described in the next section) but has resulted in some extended and unwieldy entries. As well as a tool for searching and discovery, Modes is also used to record information related to the management of the collection, including accessions, accruals, digitisation, public access and records management.

Modes can also cope with the mixed nature of the Archive and there is much advantage derived from describing all aspects of the archival holdings in a single object data structure, with common searches, views and reports. This allows interrogation of the database across the entirety of the collection. The software also allows easy access to digital objects, images, documents, video, audio files. This combination of material discovery and immediate access to downloadable digital objects will be essential components to translate to a web-based environment. The new Modes software is XML-based, facilitating the development of a web portal and improved data-sharing. A future aim will be to establish a one-click data exchange with systems such as the Archives Hub – a discovery portal bringing together descriptions of UK's archive collections from over 300 institutions.

Web access

Creating the portal

The project is now at the web development stage, which is being undertaken by the Web & Apps team. The software selected for the public-facing interface is CKAN, which is an open-source data cataloguing system. This is used by many of open government data initiatives around the world, including data.gov to promote discovery and access of open data to the public. CKAN provides a highly customisable environment and allows flexibility in the setup and implementation of the system. This has enabled ease of mapping elements across from Modes. The resulting interface will provide discovery-level catalogue descriptions and access to images and digital versions of documents and films. The hierarchical structure of the archive will also be displayed and navigable.

Web access to the archive is the first strand of a wider project to create a CKAN-based discovery portal that will provide access to polar collections across BAS. This will primarily include datasets held by the Polar Data Centre (PDC). As the UK's national repository for polar data, the PDC has a dual obligation under the Antarctic Treaty to make the results of research freely available. Additionally, as a publicly funded organisation, funded by the Natural Environment Research Council (NERC), to provide open access to its data. This includes harvesting of its metadata records by the Antarctic Master Directory and NERC Data Catalogue Service.

The BAS Discovery Metadata System (DMS) is currently the primary method of storing and distributing discovery-level metadata on datasets held by the PDC. This system was built in-house in 2004 when there was not the range of alternatives on the market that currently exist. It has been in need of updating, presenting the opportunity to take advantage of openly available technology solutions, such as CKAN, that are now available. Combining the collections PDC and Archives through the same web portal will allow improved connections between the archival material and datasets. For example, atmospheric data held by the PDC could link to data held by Archives but collected using different methods. It is still the same type of data, whether recorded on log sheets or on automatic logging systems, and so long-term data records are joined up. Published datasets could also be linked to published papers and so build a more cohesive picture of research and, with the increasing use of data DOIs, improve tracking of scientific impact and output.

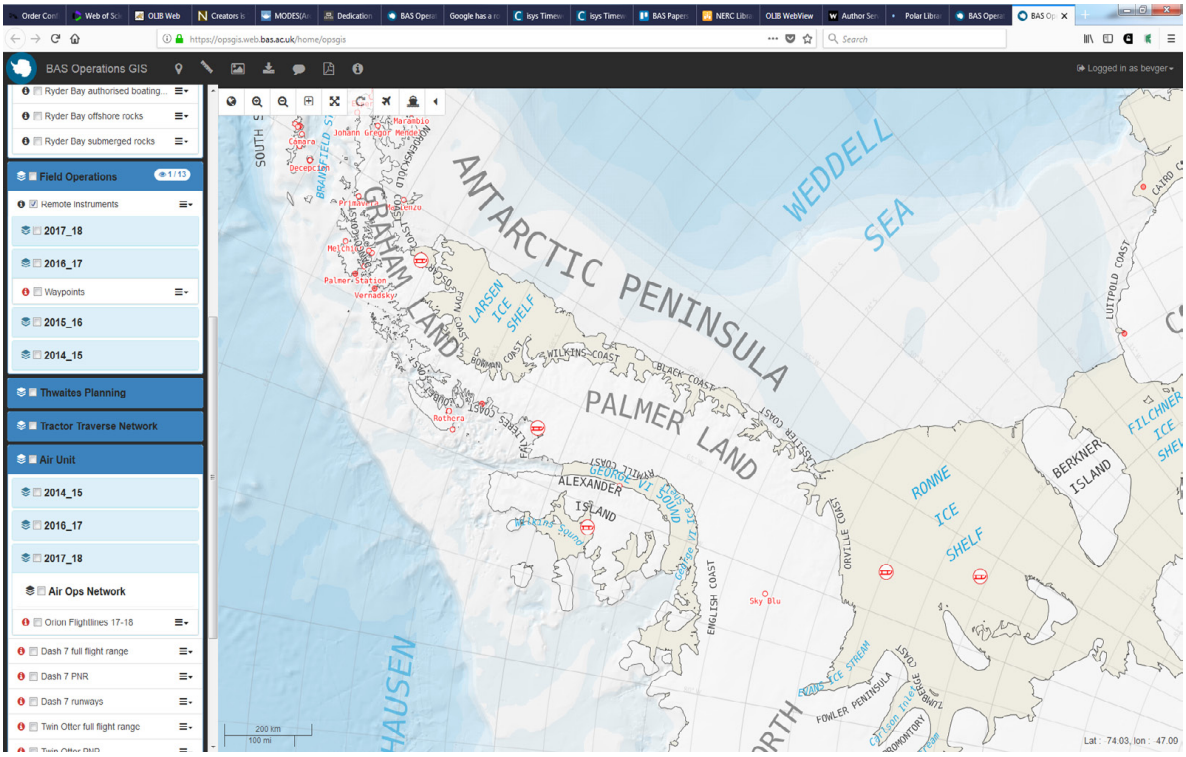
Current use cases: establishing relevance

The completion of the Archives web portal will mark the first time that the archives database has been publicly exposed. Currently, the only means of accessing the collections is through the internal catalogue. Although it is to be envisaged that bringing the Archives to a wider audience will increase public awareness of the collection, a key objective is to increase internal applications of information held within archival material.

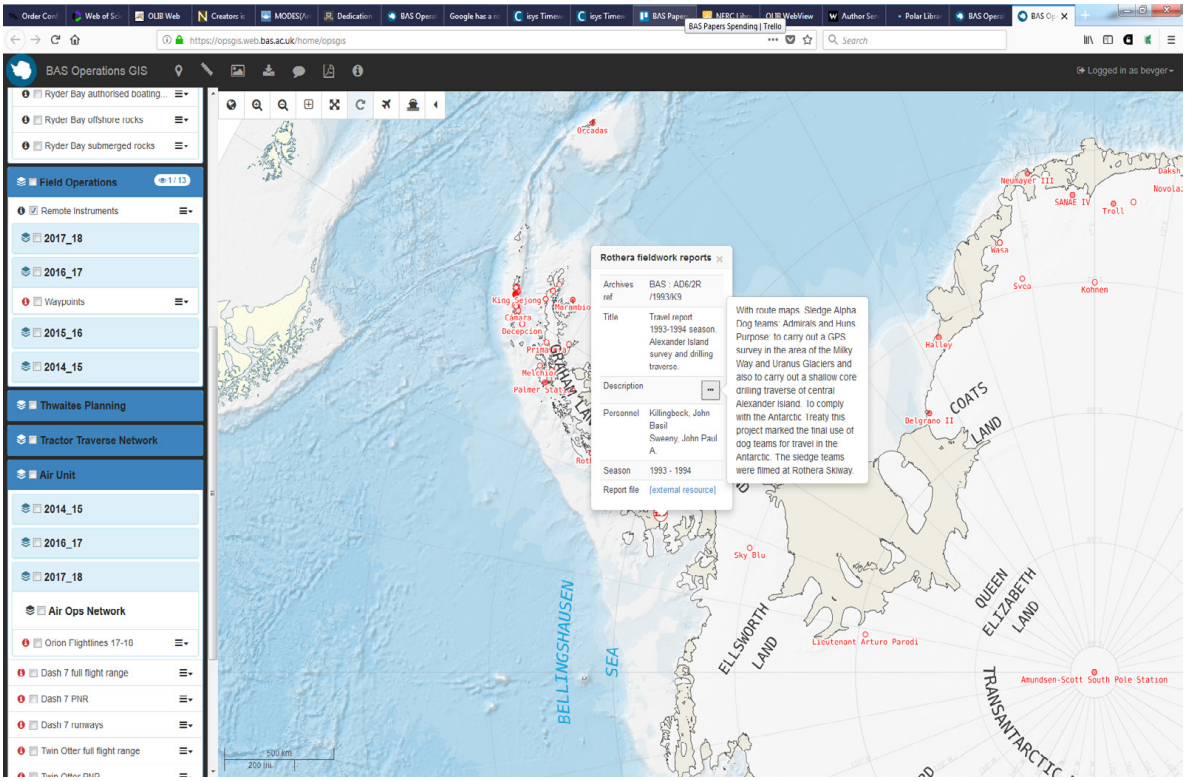
An important reason for increasing internal use of archival material is to inform current operations and science – and so benefit from past work and avoid spending resources on recreating or duplicating data collection. A recent project that came about as a response to a specific need, also acted as a case study to demonstrate the benefit of archival material for informing current operational planning. This project was carried out in collaboration with BAS's MAGIC team (Mapping and Geographic Information Centre) using the travel reports from the Archives. The travel reports are a subseries within the Base & Field reports series, which is a key series documenting the activities of the organisation. The travel reports are often written by the field general assistant or field guide in a party and detail sledge journey and field

party work. They can consist of little more than a daily account of events, and often contain a summary of the aims and objectives of the journey, list of participants, calendar of main events, sketch maps of routes and campsites. Some reports also contain statistics concerning distances travelled, sledge loads and dog performance, details of travel rations, depot contents and weather conditions. There are also comments on routes, equipment and clothing.

For this project all travel reports for a current BAS station, Rothera Station on Adelaide Island, were digitised – approximately 500-600 reports have been generated from when the station opened in 1975. Locations within the reports were georeferenced by MAGIC and added to the internal Operations GIS system, used for logistics and planning. Other key metadata was added from the archives database to improve searching, including dates, participants, summary descriptions, along with a link to download the report in full. This was an initial proof of concept but was enough to show the benefit when planning fieldwork of knowing if an area or site had been visited previously, when, by whom, the purpose and outcome etc. This project also provided a 'positive feedback mechanism' for the Archives as the georeferencing work enabled archival location data to be refined and boundary boxes or polygons of geographic areas created. With available resource, there could be plenty of scope to develop this further – for example, mapping routes of sledge journeys taken.

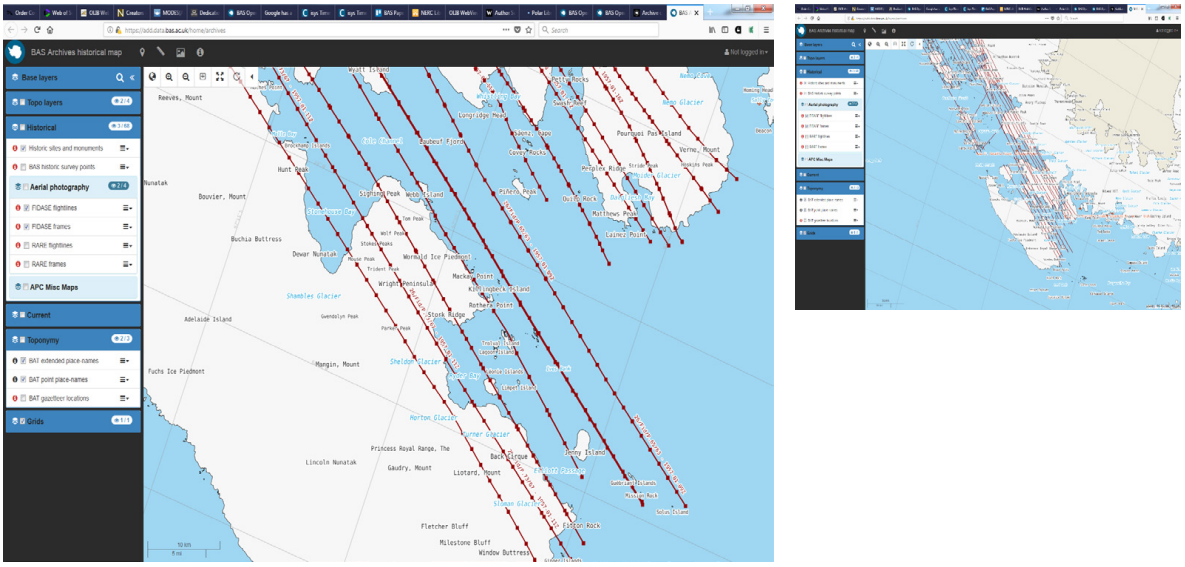


Screenshot showing the results of a participant search for "Sweeny" – the red sledge points on the Peninsula mark travel reports relating to the participant.



Screenshot showing the contextual metadata provided when clicking on one of the sledge pointer. This information is taken from the archives database and demonstrates the benefit of cataloguing to a high-level of granularity.

A similar example can be seen in the mapping of flight-lines carried out for the Falkland Islands Dependencies Survey Aerial Expedition (FIDASE), 1955-1957. This also benefitted the Archives by creating a tool for easier identification of aerial photographs held in the archives.



Screenshot showing a segment of FIDASE flight lines zoomed in – each red dot on the line provides the frame number, aiding in the location of the relevant aerial photograph within the Archives. (add.data.bas.ac.uk/home/archives)

Another data visualisation project being undertaken is the Marine Metadata project. The Marine Metadata team in the PDC compiled a comprehensive database of BAS cruise information using BAS archive records to fill information gaps. Memoirs, administrative records, Chief Officers' Deck Logs and Masters' Voyage Reports were all extensively consulted. With the help of these records, a timeline was created for the RRS *James Clark Ross* stretching back to 1991 and including its most recent of cruises, producing a daily account of where the ship was, and activities over a twenty-year period.

The travel reports and other documentation in the archives have provided information to underpin conservation and heritage work carried out by the UK Antarctic Heritage Trust (UKAHT) and the BAS Environment Office. Under the terms of the Antarctic Treaty, BAS has a responsibility to manage all of its abandoned stations and refuges. This includes the designation of those suitable as Historic Sites and their subsequent maintenance (managed under a Memorandum of Understanding with UKAHT) as well as the clean-up or removal of others. These conservation projects have been assisted by records in the Archive, such as the original construction reports and travel reports that show where depots have been left. Locations can then be georeferenced and satellites tasked to look for what remains. This is an example of repurposing archival material, without which there wouldn't be the information required to narrow down locations.

Data re-purposing and re-use – particularly re-using a certain type of data in a different way is another route into the Archives that can be developed further. For example, data on the location of snow petrel nests within archival material has been used to identify sites where scientists can analyse avian stomach oil deposits on rock – the dating of which has been used as a new method to provide a minimum age for local ice thinning. This type of research has been used along to determine ice sheet thickness changes in some parts of Antarctica. Usage of this kind shows that it's not always obvious how archival sources will be useful without specialist scientific interpretation. However, exposing the material for re-use is a starting point for starting the dialogue between the Archives and Science and so build collaborative networks of polar data.

Conclusion

In conclusion, there is still much to do but opening up the Archives brings a lot of scope and potential for future development. Linking data held within the Archives and PDC enables networks of polar data and information to be created, increasing data exchange and scope for reuse. It marks a transition from a flat catalogue to an interactive database that will fully expose the richness and diversity of this resource. There is an increasing need and benefit to be gained from data exchange, meshing archival content with different systems for different purposes and providing open linked data. As well as bringing the collections to new external audiences, this can be used to demonstrate the value of the Archives to internal stakeholders and so embed the collection within the organisation. Additionally it supports BAS in fulfilling its legal obligations of making data and information openly accessible. New technology has increased available tools and improved methods for the geospatial visualisation of archive material. It has also shown the benefits of consistently detailed cataloguing over a long period, enabling the repurposing of catalogue content to provide a narrative for other data access points. However, it is also a two-way interaction with better georeferencing informing the Archives and improving catalogue descriptions. It is also easy to envisage how these developments could be expanded upon in future, such as digitising and georeferencing panaround survey photographs so you could go to a point on a map and get a 360° view of the exact spot where, for example, a surveyor was plane-tabling on Signy in 1958.

References

ANTARCTIC MASTER DIRECTORY: https://gcmd.gsfc.nasa.gov
ARCHIVES MAP VIEWER: https://add.data.bas.ac.uk/home/archives
ARCHIVES HUB: https://archiveshub.jisc.ac.uk/
BAS ARCHIVES: https://www.bas.ac.uk/data/our-data/collections/archive-collections/
DATA.GOV: https://data.gov.uk/
DISCOVERY METADATA SYSTEM: https://data.bas.ac.uk/
MODES: https://www.modes.org.uk/
THE NATIONAL ARCHIVE: http://www.nationalarchives.gov.uk/
NERC DATA CATALOGUE SERVICE: https://csw-nerc.ceda.ac.uk/geonetwork/srv/eng/catalog.search#/home
SCOTT POLAR RESEARCH INSTITUTE: https://www.spri.cam.ac.uk/archives/
UK ANTARCTIC HERITAGE TRUST: http://www.ukaht.org/